Statistical Method Based Algorithm for Fault Detection in *Wireless Sensor Networks (WSNs)*

DEEPAK YADAV[1] AND TUHINA SAMANTA[2]

[1] *M.Tech, Department of Information Technology, IIEST Shibpur, West Bengal, India*
[2,*] *Assistant Professor, Department of Information Technology, IIEST Shibpur, West Bengal, India*
[*] *Corresponding author: t_samanta@it.iiests.ac.in*

A key issue in the wireless sensor network applications is how to accurately detect the fault status of a node when it is working in a harsh environment. The wrong detection of nodes status can cause a lot of the damage especially when it is used for critical applications. Using distributed self-fault diagnosis (DSFD) method, faults in wireless sensor networks (WSNs) can be easily detected. In this method, each sensor node collects its neighbourhood sensor node data and uses the statistical-based method for detecting its own fault status. In this paper, we discussed various statistical-based method such as standard deviation, interquartile range, median absolute deviation (MAD), S_n and Q_n scale estimator for detection of the fault in WSNs. The result of the experiment shows that standard deviation and interquartile range fails to detect the fault, if multiple nodes are faulty, while MAD, S_n and Q_n scale estimator detects the fault even 20-30% of the nodes are faulty.

1. INTRODUCTION

Wireless Sensor Networks (WSNs) consists of thousands of tiny sensor nodes having limited battery and processing power[1]. They are placed underwater for monitoring of water quality such as pH level, conductivity and contamination level of water etc. Underground applications consist of monitoring of soil quality such as humidity, nutrient level etc. There installation on the land having many applications like monitoring of forest fire, bridge vibration etc. The sensor nodes collect the data and transmitted to its base station via. router nodes for further processing. Each wireless sensor nodes consists of a sensing unit, transmission unit, CPU and power unit. Any unit failure causes the fault in WSNs, which can be corrected only by replacing the faulty unit, this type of fault is called the hard fault[2]. The main constraint of WSNs is the energy source, it is operated with the limited energy source, once the energy is drained out, the sensor nodes is no longer any part of the network. The replacement of the energy source is not possible in all types of applications. There are the many causes, which can damage the wireless sensor networks, such as failure of any subsystems causes a hard fault which can't be repaired without replacing of that subsystems, another type of fault is a soft fault which is due to the inaccurate sensor reading, software malfunctioning etc. There are several types of fault diagnosis techniques are available. They can be broadly classified into three parts: 1. Centralized 2. Distributed and, 3. Hybrid approach[2]. In a centralized approach, all the nodes are sending their data to the base station for the detection of the fault. This method put lots of burden on the network and it is also more energy consuming method. The distributed approach divides the computation among all nodes so that energy consumption should be reduced. The hybrid approach is a combination of both methods[2]. In this paper, the Q_n scale estimator, which has a higher Gaussian efficiency (82%) is used for the fault detection in WSNs[3]. Other methods such as standard deviation (SD), interquartile range (IQR), median absolute deviation (MAD) and S_n scale estimator are also used along with Q_n scale estimator, for the comparison purpose.

2. RELATED WORKS

The fault diagnosis algorithm is classified into three categories: Centralized, Distributed and Self-diagnosis approach[1]. In the centralized approach all the sensor nodes send their data to the central node for the detection of their fault status. The central node (base station), computes the fault status of all the nodes and broadcast the fault status to all nodes in the network. There are several disadvantages of this type of approach, the central node should be ultra-reliable, high computation capability and large storage are required[1]. The broadcast of fault status required multihop communication, which depletes the energy of the network quickly. The failure of the base station causes unable to detect the fault in the nodes. There are the advantage of this type of approach is that the detection latency is very high.

These disadvantages are overcome in the distributed fault diagnosis approach , where each sensor nodes participating in the diagnosis process but the final fault status is decided by the base station. In this approach, each node acts as a tester as well as a testing node. Each node also tests the fault status of neighbours node and collect all the test result known as a syndrome and send it to the base station. The base station uses the syndrome analysis approach to detect the final fault status and sends the status to all nodes[2]. This type of approach is suitable for unconstrained based networks.

The centralized and distributed approach is costly in terms of message complexity, leads to low network life due to more power consumption. The self-diagnosis approach solves the problem, in this approach, each node checks their own fault status by collecting neighbours sensor nodes data. If the fault detected then that sensor node is not participating in the network. There are several advantages of this approach. This type of approach does not put communication, memory, bandwidth and energy overload on the network. There is no need of initiator node for diagnosis of entire network and no need to know the fault status of neighbour nodes (which is the essential condition in the distributed approach), which saves the time.[1]

Lee and Choi [4] proposed neighbour co-ordination method in which sensor data compare with neighbour data at any time t and store it to the memory. This method is repeated for C times, in the final step each sensor analyzing data stored in the memory. The disadvantage of this method is that it collects the data C times (more energy loss during transmission in comparison to processing).

Liang et al.[5] proposed the statistical-based approaches which uses a weighted median based fault detection technique. In this technique, normalized data calculated which is equal to the ratio of the difference of sensed data and estimated data to the sensed data. This normalized value is compared with the threshold value, if it is greater than the threshold then it treated as the faulty node.

In three sigma edit test [1], each node collects data from the neighbour and send its data to the neighbour also and then calculate the fault status of its own node as well as neighbours node by using three sigma edit test rule.
Modified three sigma edit test rule proposed by [1], in this method, no need to send the fault status to the neighbour node. Therefore, the number of message exchange reduced. They replace the mean with median and standard deviation with normalized MAD.
Our proposed algorithm is based on the most robust scale estimator (Q_n scale estimator)[3]. There are several other scale estimators such as median absolute deviation (MAD), interquartile range (IQR), S_n scale estimator, Hodges Lehmann estimator, etc. are present but among them, Q_n scale estimator performance is more satisfactory. The Q_n scale estimator having Gaussian efficiency 82%, a more efficient estimator needs fewer observations than the less efficient one to achieve a given performance. Financial companies now using these estimators on daily basis in analysis of the behaviour of stocks. The detailed analysis of Q_n scale estimator is given in the subsequent section[6].

3. SYSTEM MODEL

The system model is consists of a network model and fault model. In the network model, the placement of sensor nodes and their way of communication are described and the fault model described how the sensor nodes get faulty during their operation.

A. Assumption, Notation and Symbols

- The energy level of each node during installation are equal and nodes are homogenous in nature.
- Each node can send packets consists of node ID and Data with its neighbours and can receive from its neighbours.
- If any nodes receive packets in which data part is missing from the packets then it considered that the sender node having faulty.
- All the nodes in the network is static and synchronized by the network protocol.
- Each node is powered with a battery source and energy loss during each packet transmission is the same for each node in the network.
- The communication link between node is assumed to be fault free.
- Each node can sense the data in a periodically manner and two nodes can communicate with each other using UDP/IP protocol.

Notation	Meaning
S	Set of all sensor nodes
s_i	i_{th} sensor node
k	k^{th} sample of data
N_{eg_i}	Set of neighbouring nodes of s_i
x_i(t)	Sensed data of s_i at time t
Nx_i	Neighbors sensed data
NT_i	Neighboring table stored at s_i
FS_i	Fault status of s_i
θ_1	lower bound of threshold
θ_2	upper bound of threshold
$FSN_{eg_{l,m}}$	Final fault status of neighbours nodes
AD_i	Absolute deviation around median
med_i	median over neighboring nodes data Nx_i

Table S1. List of important notation and their meaning.

B. Network Model

All the sensor nodes are placed in a random manner in L×L area with maximum communication range is d meter in all the direction. $S = s_i$, where $i \in [1, N]$ is the set of all sensor nodes present in the network, where N is the number of sensor nodes present in the network and N_{eg_i} is the i_{th} sensor node neighbours. $x_i(t)_{t\in[1,k]}$ is the reading of s_i node at time t sec. $x_i(t)$ can be anything based on the application such as temperature, light, noise, humidity etc. θ_1 and θ_2 are the lower and upper bound of the predefined threshold level. If the sensed data lies beyond the threshold level then it can be either an event or faulty data. IEEE 802.15.4 is used as a MAC layer protocol for communicating to neighbours node[1].

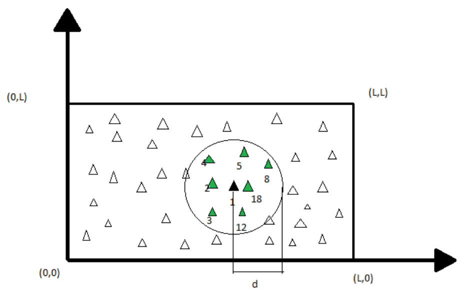

Fig. S1. Network model. The sensor nodes are represented by triangle. The green trianlge shows neighbors of black node.

C. Fault Model

Each sensor nodes collects k samples of data periodically from the environment. The data can be temperature, pressure, humidity, light etc. if the samples of data lie beyond the predefined threshold level (θ_1 and θ_2) then it can be either faulty or event data. if any data among k samples does not lie beyond the threshold level then all k samples are treated as a transient fault which can be ignored by the sensor nodes. To distinguish between faulty or event data, sensor nodes apply statistical-based method on data set collected from neighbours sensor nodes (N_{eg_i}) to verify their original status. The links between nodes are assumed to be fault-free.

4. PROPOSED METHOD

A. Distributed self-fault diagnosis method:

The proposed method is divided into two parts: 1) the temporal correlation and 2) spatial correlation. Sensor node readings are almost the same for a short interval of time, hence it is easy to find the fault in the nodes using present and past reading of the sensor node. this type of correlation is called temporal correlation. On the other hand, there is also a strong correlation between the node data which are placed geographically closer to each other. This type of correlation is known as spatial correlation. using spatial correlation nodes can compare their reading with each other to detect the faulty nodes in the network[7].

Neighbours node(N_{eg_i})	Sensor data(Nx_i)	Final status($FSN_{eg_{j,m}}$)
4	11	1
8	25	0
.	.	.
.	.	.
20	27.1	0

Table S2. Neighboring table NT_i.

B. Algorithm

Result: Fault status (FS_i) of s_i nodes.
Result: Data: $x_i(1), x_i(2), \ldots\ldots, x_i(k)$ Transmission range (Tr)
Initialization: $S = s_i$, where $i \in [1, N]$, is the set of all nodes and N_{eg_i} is i^{th} node neighbour

/*temporal correlation*/
for $s_i \in S$ **do**
  $X_i \longleftarrow x_i(1), x_i(2), \ldots\ldots, x_i(k)$
  if $x_i(1)\&\&x_i(2)\&\&...\&\&x_i(k) \le \theta_1 or \ge \theta_2$ **then**
    Collect k^{th} sample from neighbours nodes (Neg_i)
    construct neighbourhood table NT_i
    $Nx_i = x_j(k)_{j\in N_{eg_i}}$
    $Nx_i = x_i(k) \bigcup Nx_i$

    /*spatial correlation*/
    sort(Nx_i)
    /*Step 1: calculation of median(med_i) of Nx_i*/
    if $[|N_{egi}| + 1]\%2 == 0$ **then**
      med_i={Nx_i[$(|N_{egi}|+1)/2$]+ Nx_i[$(|N_{egi}|+1)/2$+1]}/2
    else
      med_i=Nx_i[$(|N_{egi}|+1)/2$]
    end
    /*Step 2: calculation of absolute deviation AD_j*/
    for $j \leftarrow 1 to (|N_{egi}|+1)$ **do**
      AD_j=(Nx_i[j]-med_i)
    end
    /*Step 3: Calculation of Q_n scale estimator($c * |x_i - x_j; i < j|$)*/
    for $j \leftarrow 1 to |N_{eg_i}|$ **do**
      for $m \leftarrow j+1 to (|N_{egi}|+1)$ **do**
        $Q_{n_{j,m}} = C_n * |Nx_i[j] - Nx_i[m]|$
      end
    end
    Convert $Q_{n_{j,m}}$ into one-dimensional array
    Sort($Q_{n_{j,m}}$)

    /*Step 4: Calculation of k_{th} order statistics*/
    $h = [N_{eg_i}/2] + 1$
    $k = h * (h-1)/2$
    $\sigma_{Q_n} = Q_{n_{j,m}}[k]$
    /*Step 5: Calculation of the fault-status of parent node (s_i)*/
    $z_i = (x_i(k) - med_i)/\sigma_{Q_n}$
    if $z_i < c * \sigma_{Q_n}$ **then**
      FS_i=0 (fault-free)
    else
      FS_i=1 (faulty)
    end
    /*Step 6: Calculation of fault status of neighbours node (N_{eg_i})*/
    for $j \leftarrow 1 to |N_{egi}| + 1$ **do**
      $z_j = AD_j/\sigma_{Q_n}$
      if $z_j < c * \sigma_{Q_n}$ **then**
        $FSN_{eg_{i,j}} = 0$ (fault-free)
      else
        $FSN_{eg_{i,j}} = 1$ (faulty)
      end
    end
  else
    /*do nothing (transient-fault occured)*/.
  end
end

C. Explanation:

Each sensor nodes periodically collect k samples, $x_i(t)_{t\in[1,k]}$ from the environment. After collection of k samples, every node test whether all the k samples, $x_i(t)_{t\in[1,k]}$ lies beyond the threshold level (θ_1 and θ_2) or not. if any single sample, $x_i(t)$ lies between the threshold level (θ_1 and θ_2), then node again collect k samples and recheck it. This avoid the transient fault in the node which occurs for small instant of time . If all k samples, $x_i(t)_{t\in[1,k]}$ are lies beyond the threshold level (θ_1 and θ_2) ,then it gives an asurity that nodes are faulty or there should be an event detected at node. Lets assume the k samples, $x_i(t)_{t\in[1,k]}$ are lies beyond the threshold level (θ_1 and θ_2),then node s_i collects neighbours k^{th} sample data and construct neighbouring table NT_i, at same time it sends data to its neighbourhood nodes N_{eg_i}. After construction of table NT_i, node s_i detect fault using statistical-based method, Step:1 it calculate the median (med_i) of data set Nx_i, where Nx_i is the magnitude of nodes sensed value from environment. Step:2, absolute deviation of median(AD_j) is computed, which is given as the difference between nodes data and its median $x_j(k) - med_{i_{j\in N_{eg_i}}}$. Step: 3 Q_n scale estimator is calculated using $Q_{n_{j,m}} = C_n * |N_{x_i}[j] - N_{x_i}[m]|; j < m$, where C_n is constant factor and its value is 2.219. $Q_{n_{j,m}}$ is difference between j^{th} and m^{th} element present in N_{x_i}. where $j, m = |N_{eg_i}|$. Step: 4 To select the k^{th} element from the sorted $Q_{n_{j,m}}$ array. first $Q_{n_{j,m}}$ array convert into one dimensional array the choose k^{th} element from it. Step: 5 calculate the z-score of parent node s_i using $z_i = (x_i(k) - med_i)/\sigma_{Q_n}$. if z_i is greater than $c * \sigma_{Q_n}$ then the fault status (FS_i) of s_i node become 1. Step:6 neighbours node fault status also calculated using $z_j = AD_j/\sigma_{Q_n}$. where $j \in N_{eg_i}$ and AD_j is absolute deviation which is calculated using $AD_j = (N_{x_i}[j] - med_i)$. where $j \in N_{eg_i}$. if z_j is greater than $c * \sigma_{Q_n}$ then the $j^{th} neighbournode \in N_{eg_i}$ become faulty node.

D. Analysis of statistical method

the sensor nodes collects k samples during t time. these k samples are treated as single frame of data. it is given as: $x_i(1), x_i(2), x_i(3), ..., x_i(k)$. if all these samples lies beyond the threshold level (θ_1 and θ_2) then the k^{th} sample of all the neighbours $x_j(k)_{j\in N_{eg_i}}$ are collected and used for fault detection. Before apply statistical method, it is assumed that these data set $x_j(k)_{j\in N_{eg_i}}$ are follows the normal distribution $N(A, \sigma_j^2)$. where A is actual value of sensor reading and σ_j is standard deviation of data set. Any sensor reading can be written as: $x_j(k) = A + \varepsilon_j(k)$, where A represent the actual data and $\varepsilon_j(k)$ is the fault in the data[1]. The fault status of the node is calculated using Q_n estimator. A good estimator having a high breakdown point and high Gaussian efficiency[3]. The breakdown point of Q_n estimator is 50% and Gaussian efficiency is 82%[3]. Q_n estimator is defined as: $\sigma_{Q_n} = C * (|x_i - x_j|; i < j)_k$, where C is the constant factor and k is the k^{th} order statistic and defined as $k = h * (h-1)/2$, where $h = [n/2] + 1$ and n is total number of difference pair, $|x_i - x_j|$. The main purpose to choose the Q_n estimator is that it attains 82% Gaussian efficiency in comparison to the other methods[3].

To distinguish between faulty and non-faulty node, z- score is used which is given as:

$$z = \frac{|x_i - median(x_i)|}{\sigma_{Q_n}(x)} \tag{S1}$$

if the value of $0 < z < 3$ then it is treated as fault free node, otherwise it is treated as faulty node. the 99.7% of the data lies within 3σ.

Statistical Methods	Breakdown Point	Efficiency	Running time and Space
Median Absolute Deviation	50%	37%	$O(n)$ and $O(n)$
S_n	50%	58%	$O(n \log n)$ and $O(n)$
Q_n	50%	82%	$O(n \log n)$ and $O(n)$

Table S3. Comparison of various statistical method[3].

5. EXPERIMENTAL ANALYSIS

This experiment performed after the nodes collect its neighbourhood node data. In this experiment, a node tries to find its fault status by collecting its neighbourhood data and at the same time, it also exchange its data to its neighbour. when it collects all the data from its neighbours then it applies the statistical method for fault detection. We tested different statistical method such as standard deviation, IQR, MAD, S_n and Q_n estimator to detect the fault in sensor nodes and found that standard deviation which is non-robust highly influenced by outliers. IQR (interquartile range) which is calculated using the difference between 75^{th} percentile and 25^{th} percentile of data set is also a not robust comparison to MAD, S_n and Q_n estimator[6].

A. Experimental Setup

This experiment is performed on the Intel i3 core CPU having 4 GB of RAM and 2.30 GHz clock speed. The nodes are randomly placed in the L X L area and assigned a value to all the nodes. After collecting data from neighbour nodes, each node tried to detect its fault using statistical method implemented on r-language. Output of all the method are discussed below:

A.1. Experiment-1

Let us assume that nodes are measuring environment temperature and the lower threshold and upper threshold are 23 °C and 29 °C. If the sensor reading lies beyond the threshold, it is treated as a faulty node or event node. The final status is given after verification by the neighbour's node data. A node senses 25°C (we only focus on magnitude), surrounded by the five neighbourhood nodes having values 25.5°C, 26°C, 26.8°C, 27°C and 27.1°C. Since these are neighbour nodes, their values are highly correlated to each other and we assumed that all nodes are fault-free. The fault status of nodes are calculated using various statistical method are given in table S4. There is some notation used in the table. FS stands for fault status, z stands for z-score. FS=1 for faulty nodes and 0 for fault-free nodes.

Data	Zs (SD)	FS (SD)	FS (IQR)	Zs (MAD)	FS (MAD)	Zs (S_n)	FS (S_n)	Zs (Q_n)	FS (Q_n)
25	1.42	0	0	1.46	0	1.47	0	0.79	0
25.5	0.844	0	0	0.94	0	0.94	0	0.51	0
26	0.268	0	0	0.42	0	0.42	0	0.22	0
26.8	0.652	0	0	0.42	0	0.42	0	0.22	0
27	0.822	0	0	0.63	0	0.63	0	0.37	0
27.1	0.99	0	0	0.73	0	0.73	0	0.39	0

Table S4. Nodes are fault-free.

A.2. Experiment-2

Parent node set to be faulty, which sensed incorrect value (10°C), which is beyond the lower threshold level. Parent node data is mentioned in first row of table S5.

Data	Zs (SD)	FS (SD)	FS (IQR)	Zs (MAD)	FS (MAD)	Zs (S_n)	FS (S_n)	Zs (Q_n)	FS (Q_n)
10	2.03	0	1	17.03	1	17.19	1	7.38	1
25.5	0.26	0	0	0.94	0	0.94	0	0.40	0
26	0.34	0	0	0.42	0	0.42	0	0.18	0
26.8	0.45	0	0	0.42	0	0.42	0	0.18	0
27	0.48	0	0	0.63	0	0.63	0	0.27	0
27.1	0.49	0	0	0.73	0	0.73	0	0.31	0

Table S5. Parent node is faulty.

A.3. Experiment-3

Single neighbour node along with parent the node is set to be faulty. parent node is mentioned in first row and neighbour node is mentioned in second row. The result shown in the table S6.

Data	Zs (SD)	FS (SD)	FS (IQR)	Zs (MAD)	FS (MAD)	Zs (S_n)	FS (S_n)	Zs (Q_n)	FS (Q_n)
10	1.35	0	0	17.03	1	15.72	1	7.38	1
11	1.23	0	0	15.99	1	14.76	1	6.93	1
26	0.55	0	0	0.42	0	0.38	0	0.18	0
26.8	0.65	0	0	0.42	0	0.38	0	0.18	0
27	0.68	0	0	0.62	0	0.57	0	0.27	0
27.1	0.69	0	0	0.73	0	0.67	0	0.31	0

Table S6. Two nodes are faulty.

A.4. Experiment-4

In this experiment multiple nodes set to be faulty. There are total three nodes out of six nodes are faulty, which is the 50% of the data is contaminated. Faulty nodes are present in first, second and sixth row of the table S7.

Data	Zs (SD)	FS (SD)	FS (IQR)	Zs (MAD)	FS (MAD)	Zs (S_n)	FS (S_n)	Zs (Q_n)	FS (Q_n)
10	1.04	0	0	1.38	0	1.15	0	0.47	0
11	0.97	0	0	1.23	0	1.08	0	0.44	0
26	0.06	0	0	0.03	0	0.028	0	0.01	0
26.8	0.11	0	0	0.03	0	0.03	0	0.11	0
27	0.13	0	0	0.05	0	0.04	0	0.17	0
50	1.71	0	1	1.99	0	1.65	0	0.67	0

Table S7. Multiple nodes are faulty.

B. Results

In the first experiment, there is no fault in any sensor nodes. Every statistical method gives correct fault status (FS) i.e. FS =0. When the fault is introduced in the second experiment, every statistical method gives correct output i.e. FS=0 except standard deviation(SD). Hence, the standard deviation (SD) method is not efficient to detect the faults. it is not able to detect the single fault in the node. When the single neighbour node became faulty, the IQR method fails to detect the fault, only MAD, S_n and Q_n estimator gives the correct output. When half of the nodes become faulty then none of the methods gives correct output. From the above result, it is clear that standard deviation (SD) and IQR methods are not the robust statistic comparison to the MAD, S_n and Q_n estimator. MAD and S_n method also become less efficient if the number of nodes increases. The Q_n estimator having higher efficiency to detect the fault.

6. CONCLUSION

Statistical-based methods are the easiest method to implement in the sensor network. this method uses the normal distribution of data set for the detection of the fault in the sensor network. In this paper, Q_n estimator statistical-based method proposed for the detection of the fault in the nodes. This method is verified by the simulation using r-language and found that its performance is better for a small degree of neighbours node as well as a large number of nodes. Other statistical methods used in this paper are standard deviation, interquartile range, median absolute deviation and S_n scale estimator. standard deviation not robust method, since it is based on a mean of data set hence a single too large fault in magnitude can influence the fault status of the other nodes. The interquartile range is better than standard deviation because it is based on the median of data set which is not influenced by extremely high or extremely low value but it is not much efficient than median absolute deviation and S_n estimator statistical-based method. for smaller node degree median absolute deviation, S_n and Q_n estimator are perform well but when the node degree increase they are not comparatively efficient than Q_n scale estimator. The Q_n estimator having higher Gaussian efficiency (82%), hence it is chosen for fault detection. this algorithm is tested on r-language and shown better performance in the detection of faulty nodes. It can detect up to 20% to 30% faulty nodes within the neighbourhood. The future work includes simulation of the algorithm on the network simulator and hardware implementation of the model.

REFERENCES

1. M. Panda and P. M. Khilar, “Distributed self fault diagnosis algorithm for large scale wireless sensor networks using modified three sigma edit test,” *Ad Hoc Networks*, vol. 25, pp. 170–184, 2015.
2. T. Muhammed and R. A. Shaikh, “An analysis of fault detection strategies in wireless sensor networks,” *Journal of Network and Computer Applications*, vol. 78, pp. 267–287, 2017.
3. M. Daszykowski, K. Kaczmarek, Y. Vander Heyden, and B. Walczak, “Robust statistics in data analysis—a review: Basic concepts,” *Chemometrics and intelligent laboratory systems*, vol. 85, no. 2, pp. 203–219, 2007.
4. M.-H. Lee and Y.-H. Choi, “Fault detection of wireless sensor networks,” *Computer Communications*, vol. 31, no. 14, pp. 3469–3475, 2008.
5. J.-L. Gao, Y.-J. Xu, and X.-W. Li, “Weighted-median based distributed fault detection for wireless sensor networks.,” *Ruan Jian Xue Bao(Journal of Software)*, vol. 18, no. 5, pp. 1208–1217, 2007.
6. P. J. Rousseeuw and C. Croux, “Alternatives to the median absolute deviation,” *Journal of the American Statistical association*, vol. 88, no. 424, pp. 1273–1283, 1993.
7. N. Peng, W. Zhang, H. Ling, Y. Zhang, and L. Zheng, “Fault-tolerant anomaly detection method in wireless sensor networks,” *Information*, vol. 9, no. 9, p. 236, 2018.